Interactive Visual Analysis Promotes Exploration of Long-Term Ecological Data

Working Group at LTER ASM 2012

Organizers: Tuan Pham (pham@eecs.oregonstate.edu)
Julia Jones

Date: Monday Sep 10, 2012 16:00 – 18:00

Location: Ruesch Auditorium – Hobbs

More Info: **http://goo.gl/9IDiJ**

HJA LTER Oregon State UNIVERSITY

Contributors:

Computer Science:	Tuan Pham Ronald Metoyer
Ecology:	Julia Jones Fred Swanson Steven Highland Jeff Miller
Information Manager:	Don Henshaw

Brief Introduction – Get to know each other

- What is your name? your position?
- How is your work related to this working group?
- What do you expect from this working group?

Goals – Information Exchange

- Discuss how ecologists formulate hypotheses and structure analyses of long-term ecological data
- Discuss how interactive visual analysis (and other tools) may help with the process
- Learn and help evaluate a visual analysis tool under development
- Lead to potential collaboration

Classic example: Anscombe's quartet

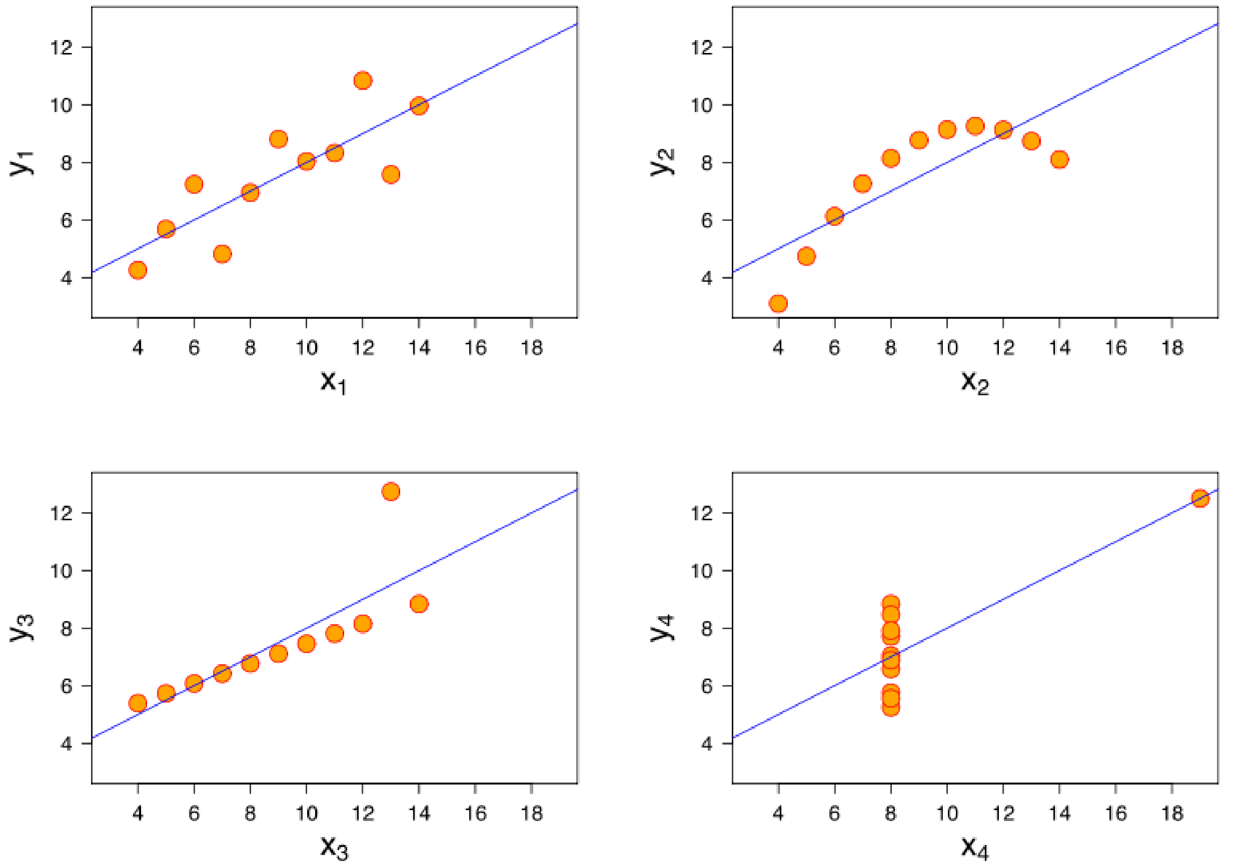
Four data sets have the same statistical properties: mean, variance, and correlation

Anscombe's quartet

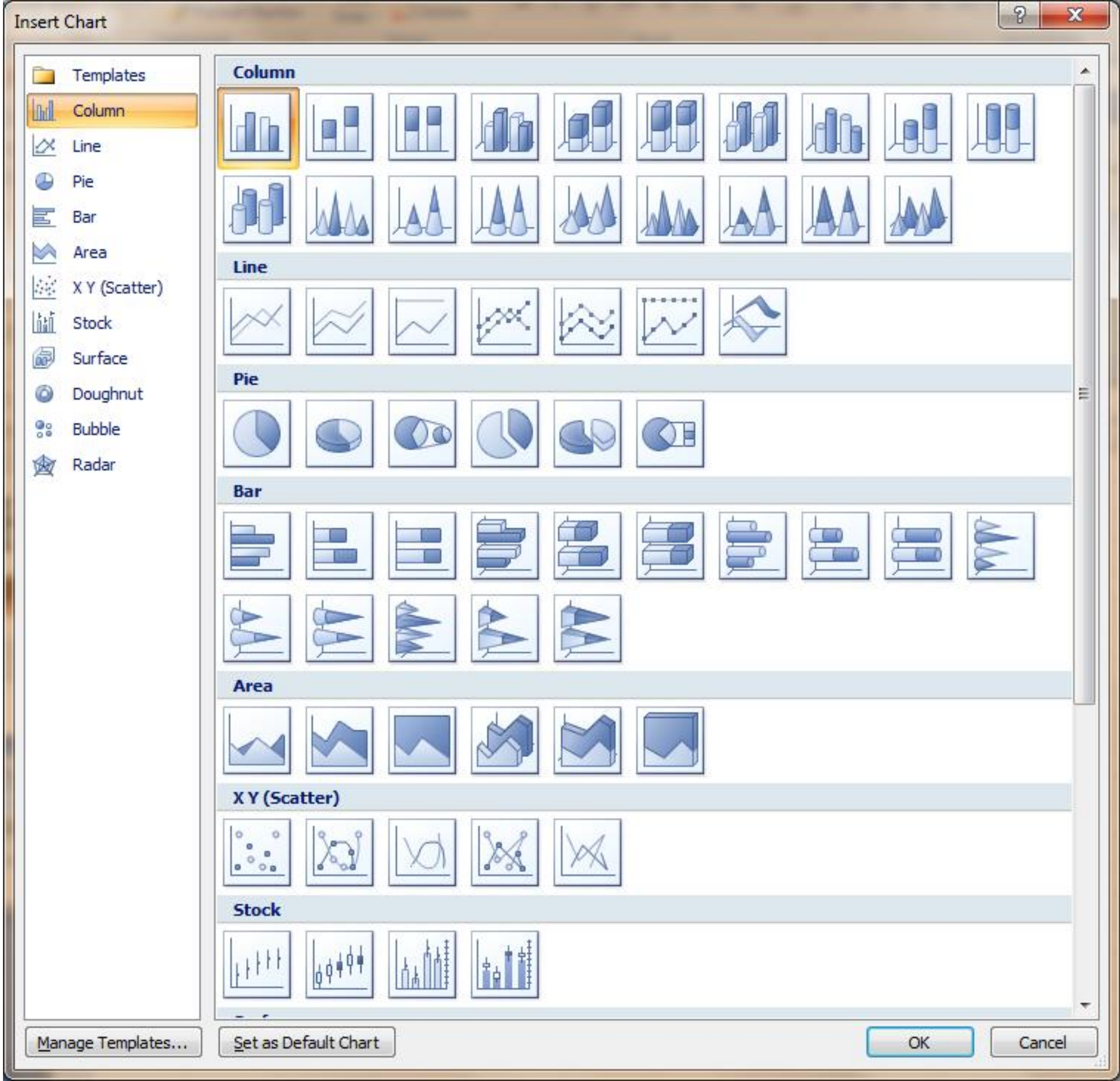
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Classic example: Anscombe's quartet

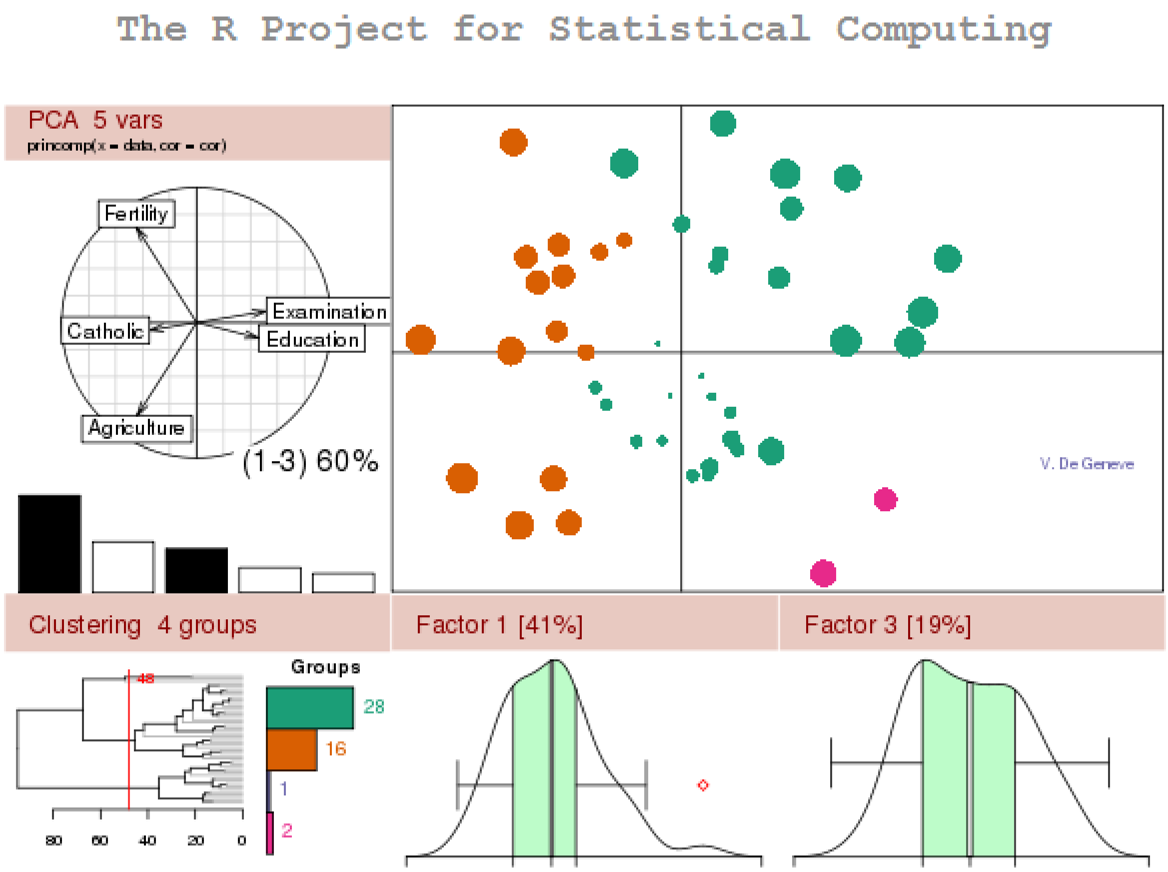
But they look different when visualized…



Charts in Excel



Charts in R


The R Project for Statistical Computing
PCA 5 vars
princomp(x = data, cor = cor)
Fertility
Examination
Catholic
Education
Agriculture
(1-3) 60%
V. De Geneve
Clustering 4 groups
Factor 1 [41%]
Factor 3 [19%]
Groups
48
28
16
1
2
80
60
40
20
0

Charts in ecological publications

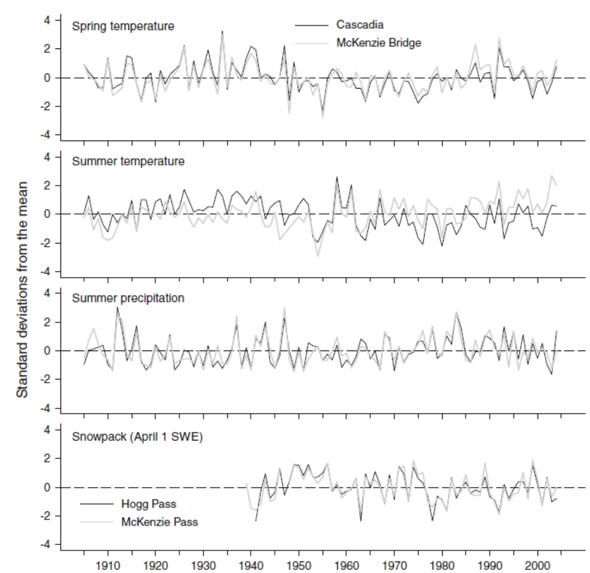
Plant Ecol
DOI 10.1007/s11258-012-0021-9

Spatio-temporal patterns of tree establishment are indicative of biotic interactions during early invasion of a montane meadow

Janine M. Rice · Charles B. Halpern · Joseph A. Antos · Julia A. Jones

Received: 27 June 2011 / Accepted: 5 January 2012

Abstract Tree invasions of grasslands are occurring globally, with profound consequences for ecosystem structure and function. We explore the spatio-temporal dynamics of tree invasion of a montane meadow in the Cascade Mountains of Oregon, where meadow loss is a conservation concern. We examine the early stages of invasion, where extrinsic and intrinsic processes can be clearly delineated. In a 0.21-ha plot, we mapped and aged 929 trees ≥0.3-m tall, yielding a detailed record of the spatio-temporal dynamics of invasion. For the examining the spatial associations of trees to a distance of 5 m and how these changed over time. We used multiple methods including uni- and bivariate forms of the Ripley's *K* and pair-correlation function (pcf) (corrected for inhomogeneity), the J-function, an evolving nearest-neighbor metric, and a test for directional bias in establishment. *Pinus* and *Abies* contributed in contrasting ways to the pace and spatial structure of invasion. Shade-intolerant *Pinus* tended to establish in the open, initiating clusters. In contrast,



Do you see any shortcomings with the charts provided by spreadsheet or statistical software?

Issues with the current charts

Standard charts provided by spreadsheets and statistical software packages

Static and non-interactive

Limited support of data exploration and collaboration among scientists

Our work aims to incorporate existing standard charts into a simple interface with **INTERACTIVITY**.

Outline for Discussion

- Background
  - DataVis = InfoVis + SciVis
  - Long-term Ecological Data
- Demos and Hands-on Exercises (60 mins)
- Discussion and Feedback (20-30 mins)

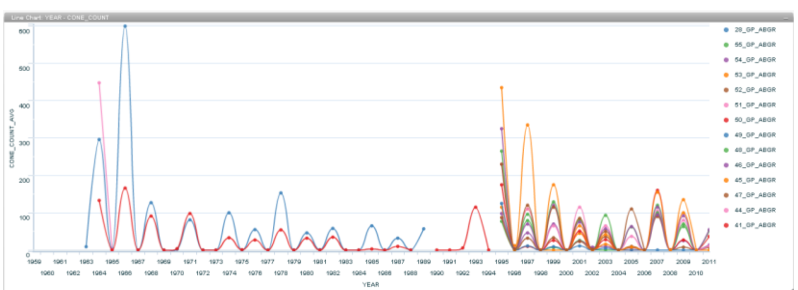
DataVis = InfoVis U SciVis

Information Visualization

Representation of **abstract data**, which do not have natural mappings to 2D or 3D space

Purposes: trends, patterns, distributions, gaps, outliers,..

Focus of this work!

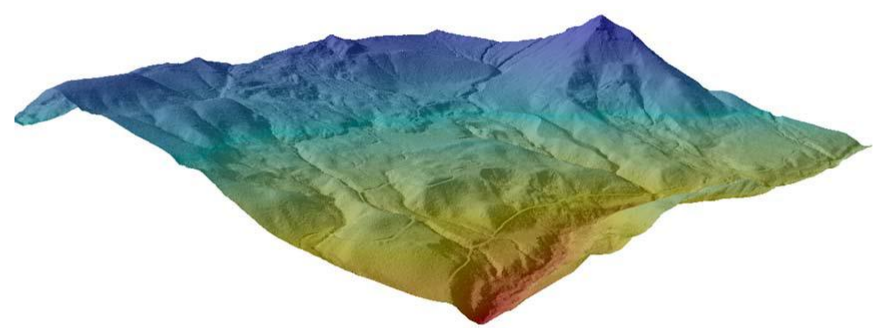


Cone production collected over time

Scientific Visualization

Data have natural mappings to 2D or 3D

Purposes: Rendering volumes or surfaces realistically



LiDAR Remote Sensing Data: HJ Andrews and Willamette National Forest, OR

The boundary between InfoVis and SciVis is fuzzy!

Long term Ecological Data

Ecological Standpoint: observational samples of living and nonliving components of ecosystems that are sampled over long-term period and described by multiple biotic and abiotic variables of varying types

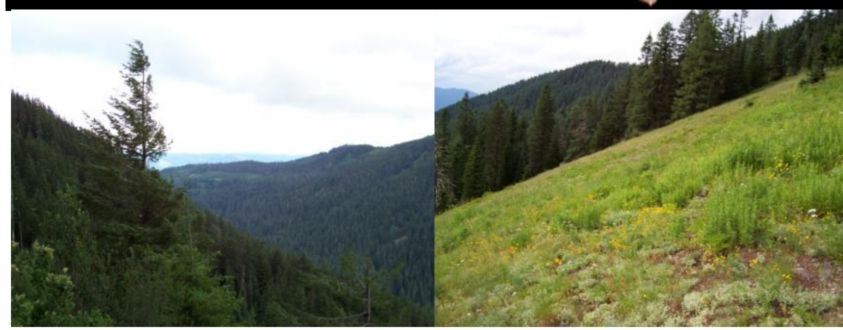
Computing Standpoint: large multivariate, geospatial, and time-series relational data.

Purposes: Gain insight into trends, patterns, distributions, gaps, outliers over long-term period

Real-world Ecological Data Sets

Moth

Vegetation

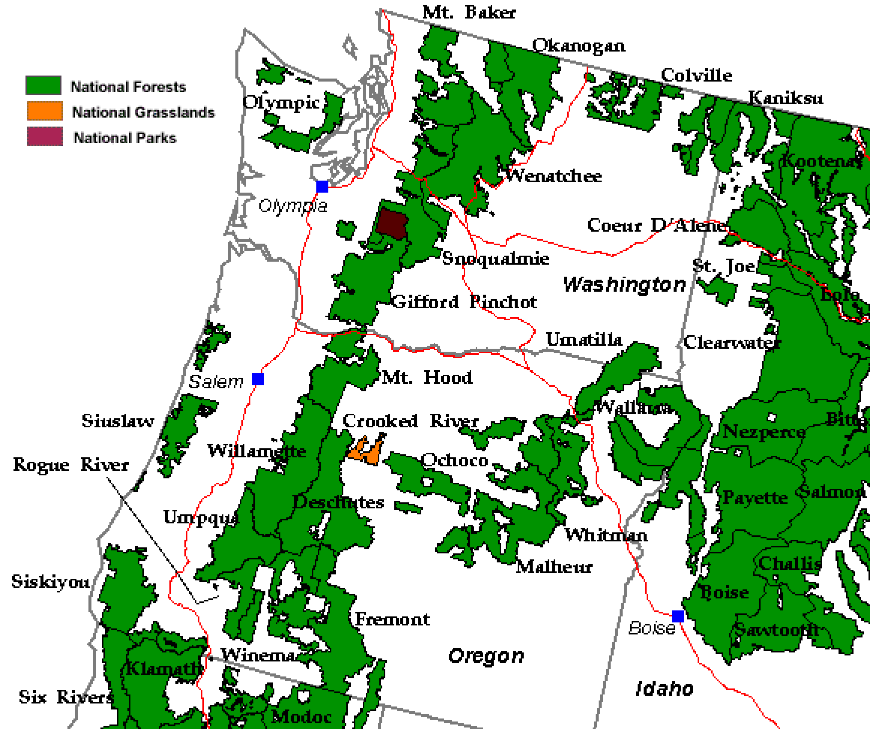


Cone Production

Cone Production Data Set: 50-year Record

Cone count observations
of upper slope conifers:

- **45,704** observations
- **934** trees of **9** species
- **10** national forests (OR-WA)
- **53** years

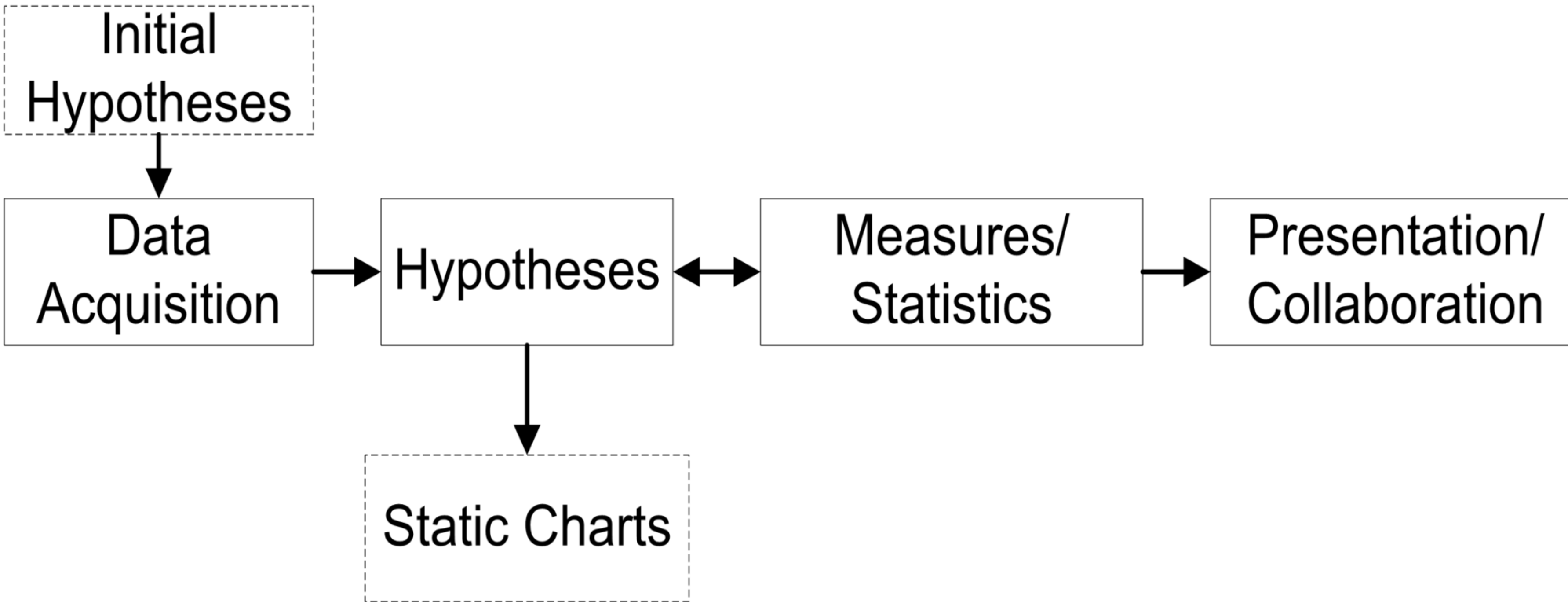


Cone Production Data Structure

Variable Name	Type	Description
SPECIES	nominal	Species code
TREE_NR	nominal	Tree number, unique for plot
FOREST	nominal/spatial	National forest code
LOCATION	nominal/spatial	Location code (within forest)
PLOT	nominal	Unique plot number (within location)
YEAR	ordinal/time-based	Sampling year
CONE_COUNT	quantitative	Number of cones
DBH	quantitative	Diameter at breast height
STATUS	nominal	Status of tree (live, dead, missing)

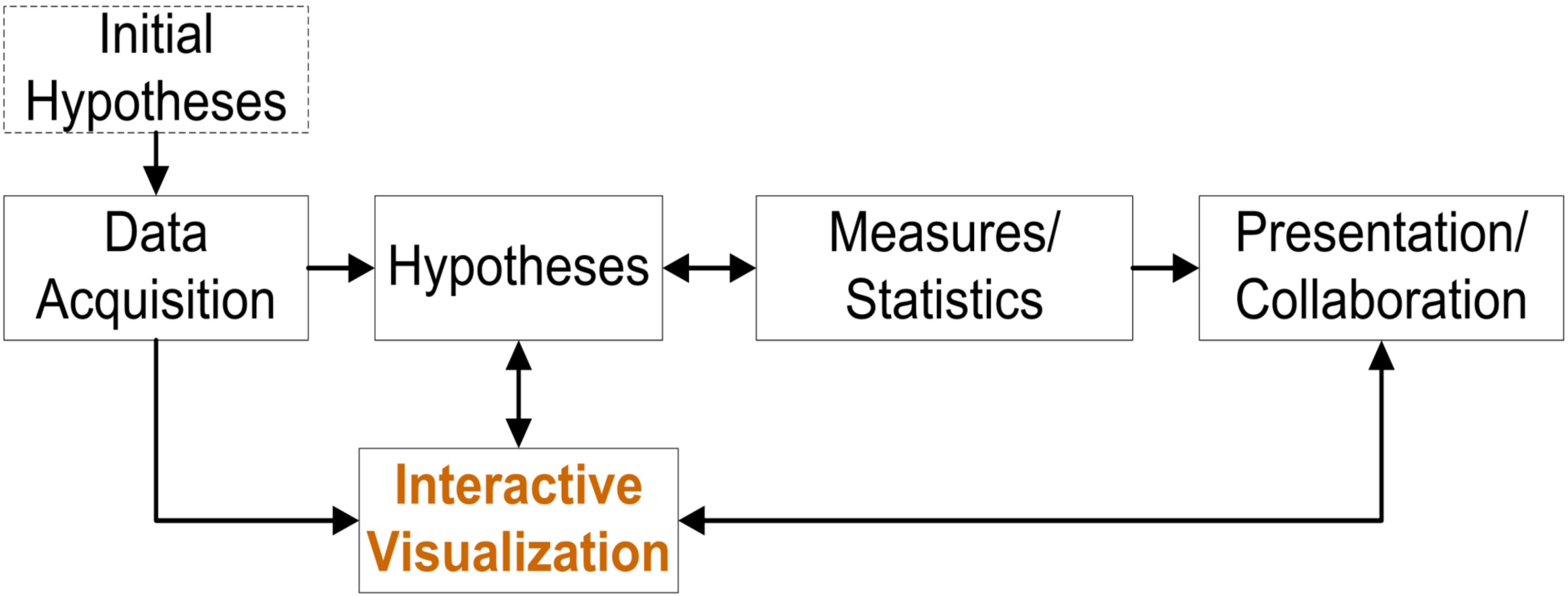
What is your approach to analyzing the cone count (or a similar) data set?

Traditional Approach to Long-term Ecological Data



Histograms, Scatter plots

Proposed Approach: Visualization Driven

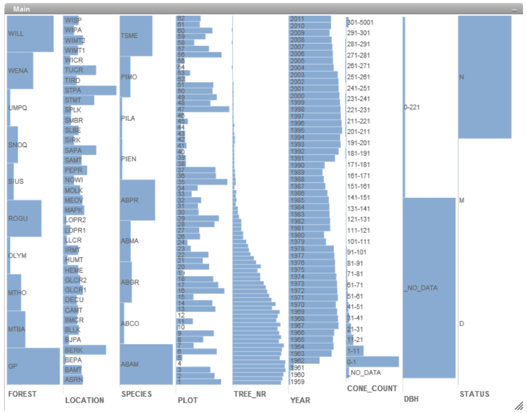


Histograms,
Scatter plots
...

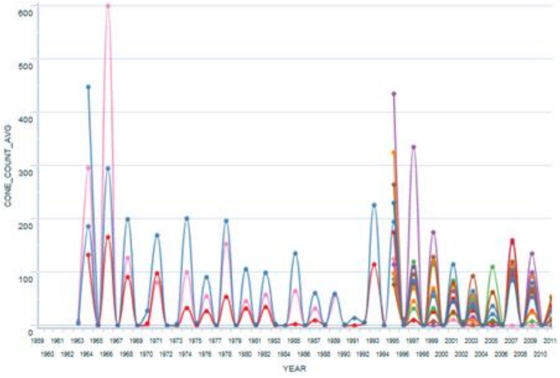
Visual Analysis = Visualization + Interaction "with user in the loop"

Cone Production Data Set: Analytical Needs

Distributions of samples

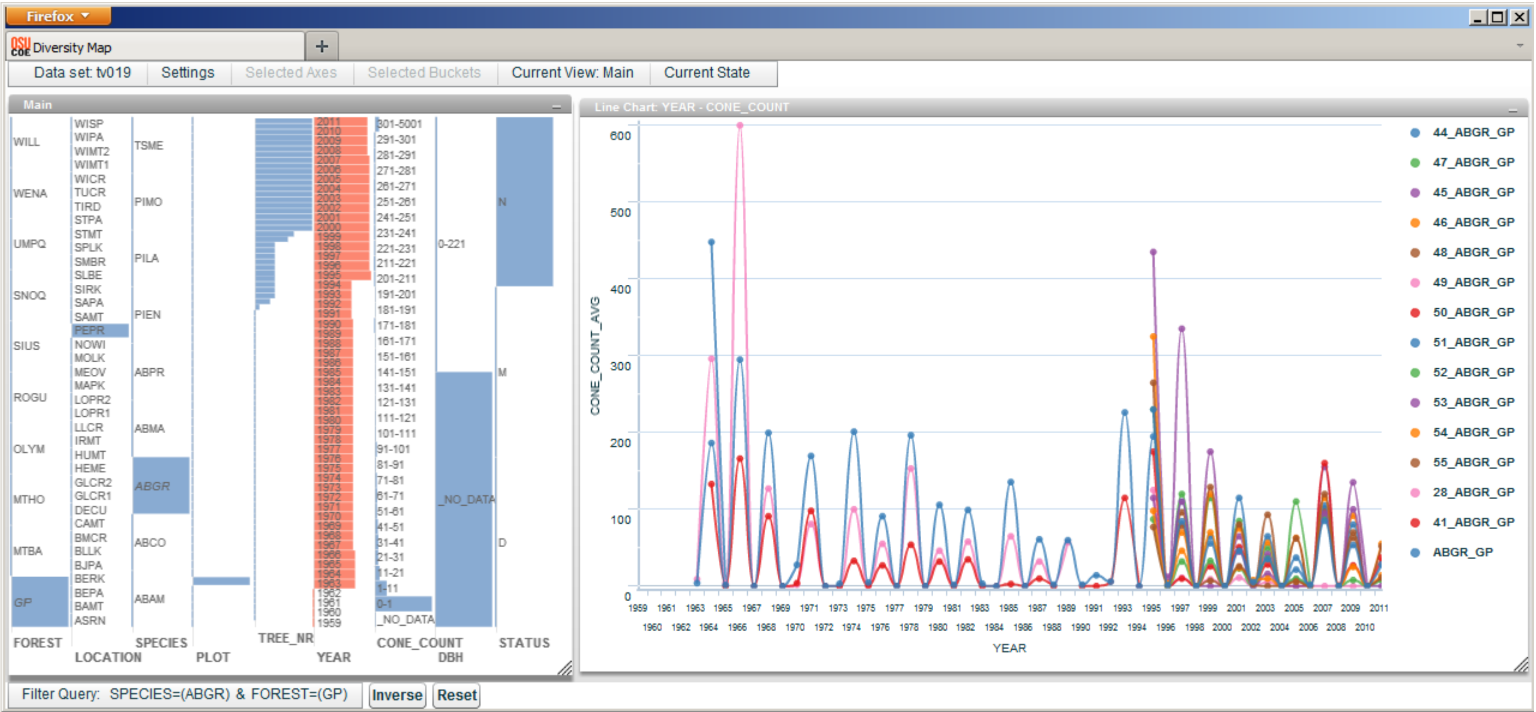


Patterns of **temporal synchronicity** of cone count



Collaborative data exploration

Visual Analysis of the Cone Production Data

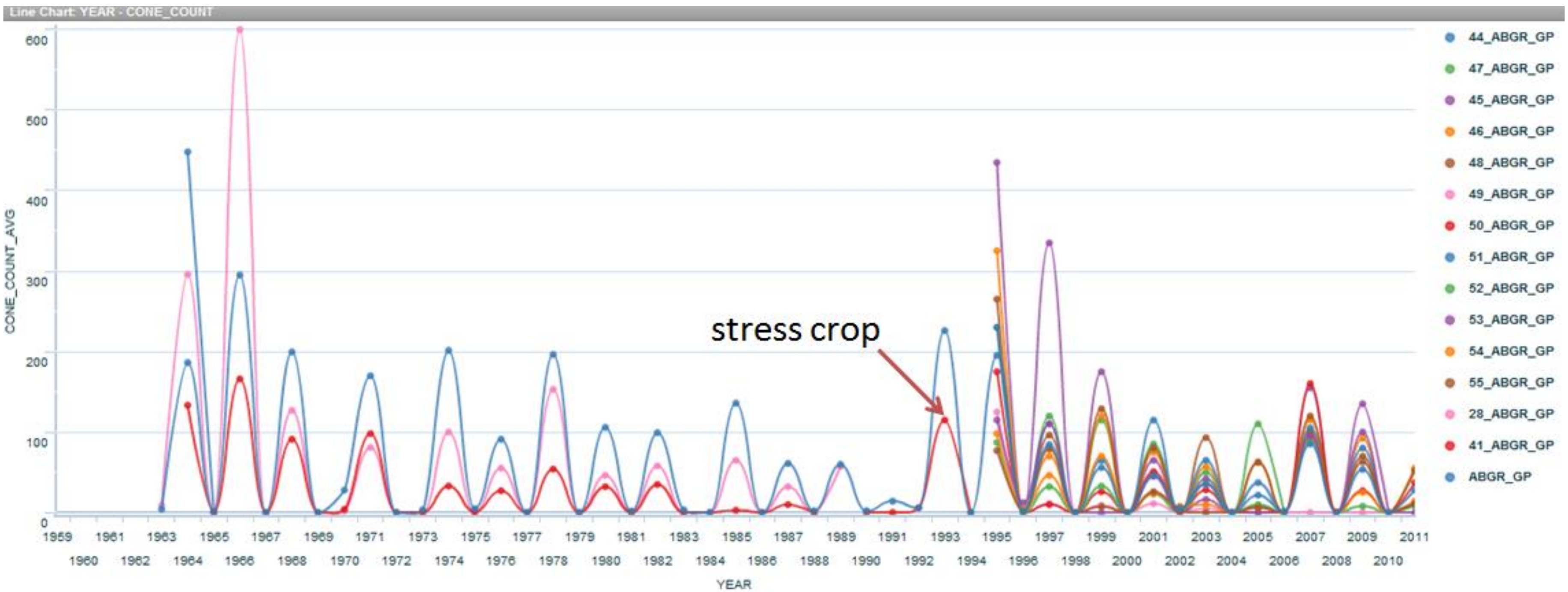


Demo: http://web.engr.oregonstate.edu/~pham/dm/cone/

T. Pham, S. Highland, R. Metoyer, D. Henshaw, J. Miller, J. Jones. **Interactive Visualization of Spatial and Temporal Patterns of Diversity and Abundance**. In *Proceedings of Environmental Information Management (EIM2011)*, 2011

T. Pham, R. Metoyer, J. Jones, and F. Swanson. **Interactive Visual Analysis of Long-Term Ecological Data**. In Preparation for Ecological Informatics Journal

Cone Production in Synchrony



URL: http://web.engr.oregonstate.edu/~pham/dm/cone/#stid=??

Hands-on Exercises

Link: **http://goo.gl/9IDiJ**

Visualization Prototype of the Cone Production Data Set

http://web.engr.oregonstate.edu/~pham/dm/cone

Visualization Prototype of the Vegetation Cover Data Set

http://web.engr.oregonstate.edu/~pham/dm/veg/

Implementation of the Tool

Web-based (Flex/Flash)

No installation required

Available everywhere

Data loaded from MySQL

Potentially loaded from any DBMS

Data import is under development

Using the visualization, did you gain any specific insights into the data sets?

Comments, Feedback on the Tool

- What aspect(s) of the tool did you like most?
- What aspect(s) of the tool did you dislike most?
- If possible, how would you change the tool to improve it?

Other related working groups

A Researcher's Guide to Automating Analysis of LTER Data

Organizer: John Porter et al.

Time: Tue, 09/11/2012 - 10:00 - 12:00

Room: Longs Peak - Diamond West

Ecological Research from a Truly Long-Term Perspective: What Needs to be Done Today to Support Ecological Analysis in 2100?

Organizer: Susan Stafford

Time: Wed, 09/12/2012 - 10:00 - 12:00

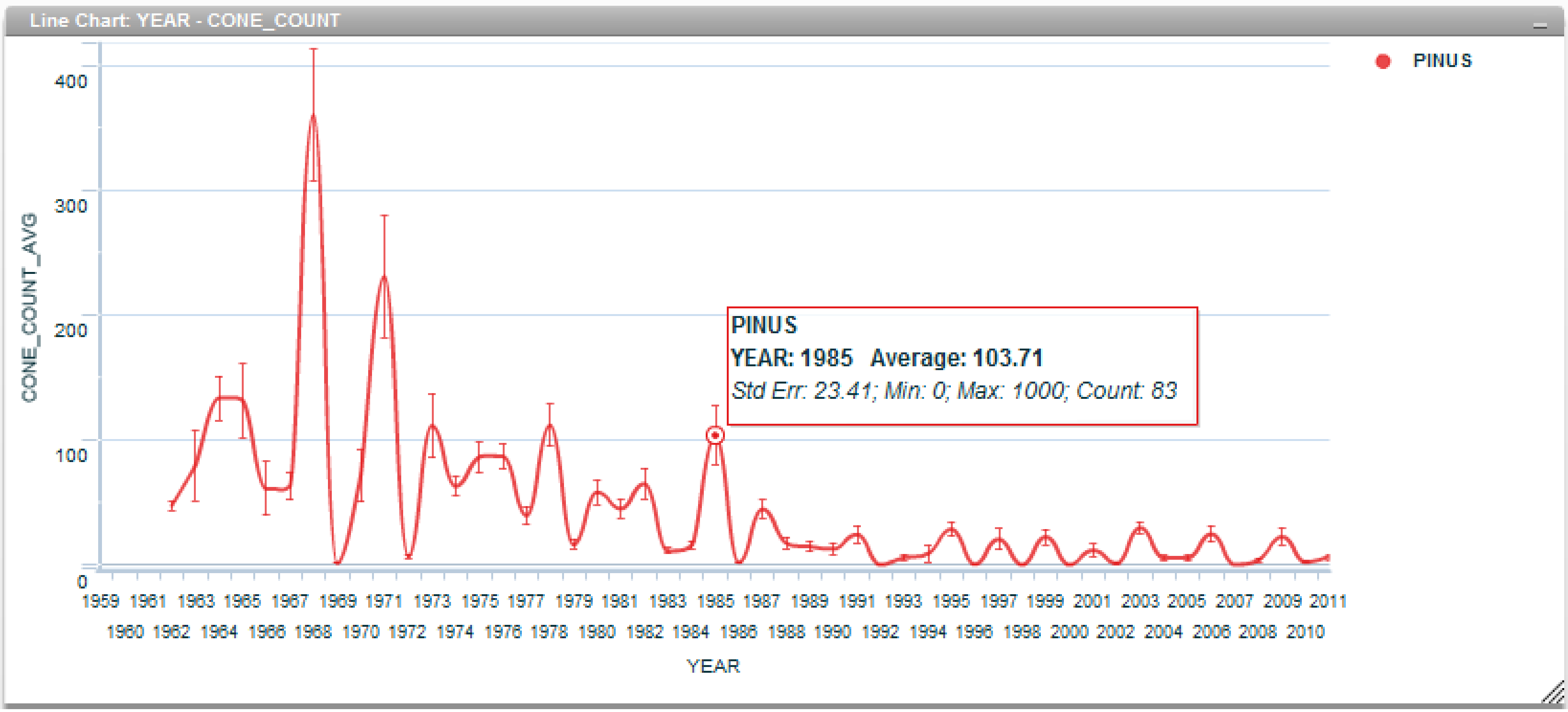
Room: Wind River A

Contributors:

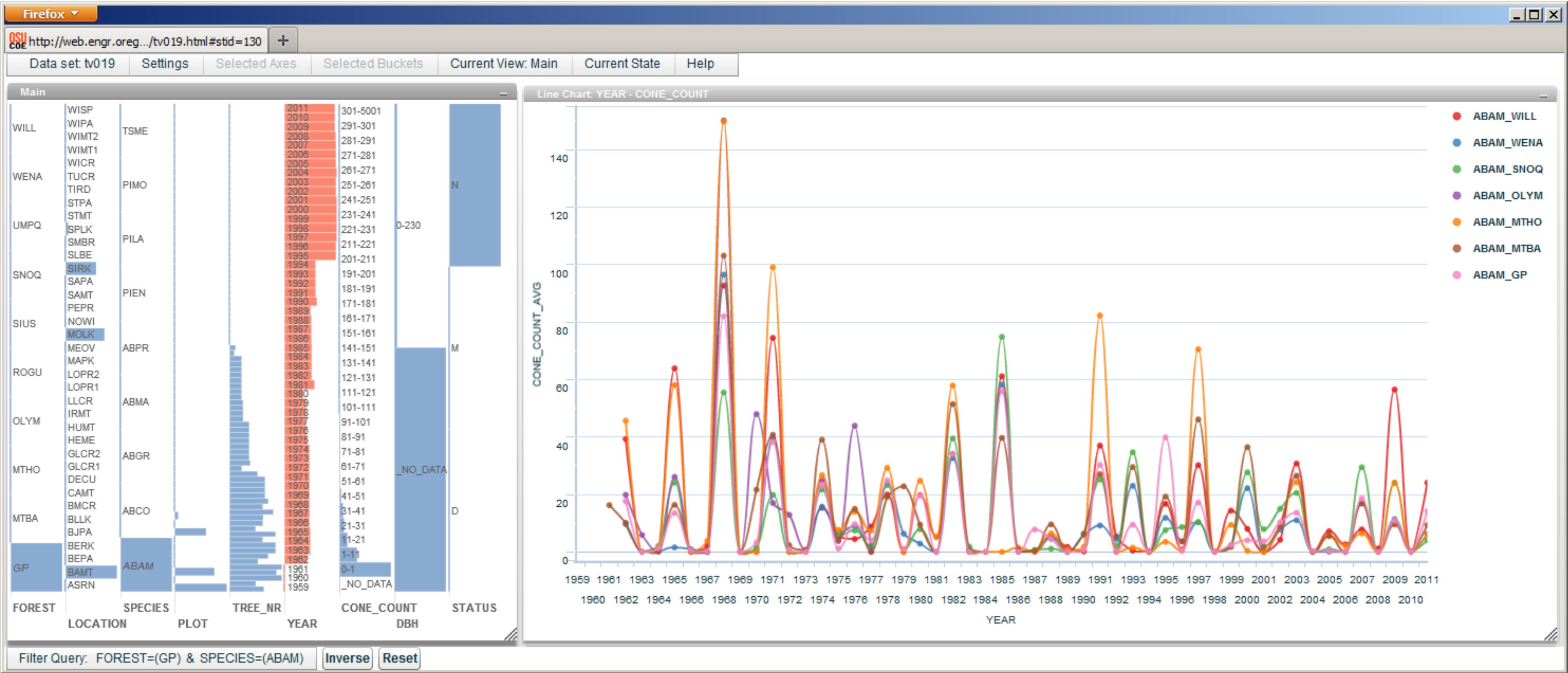
Computer Science:	Tuan Pham Ronald Metoyer
Ecology:	Julia Jones Fred Swanson Steven Highland Jeff Miller
Information Manager:	Don Henshaw

Contact: Tuan Pham pham@eecs.oregonstate.edu

Additional Slides

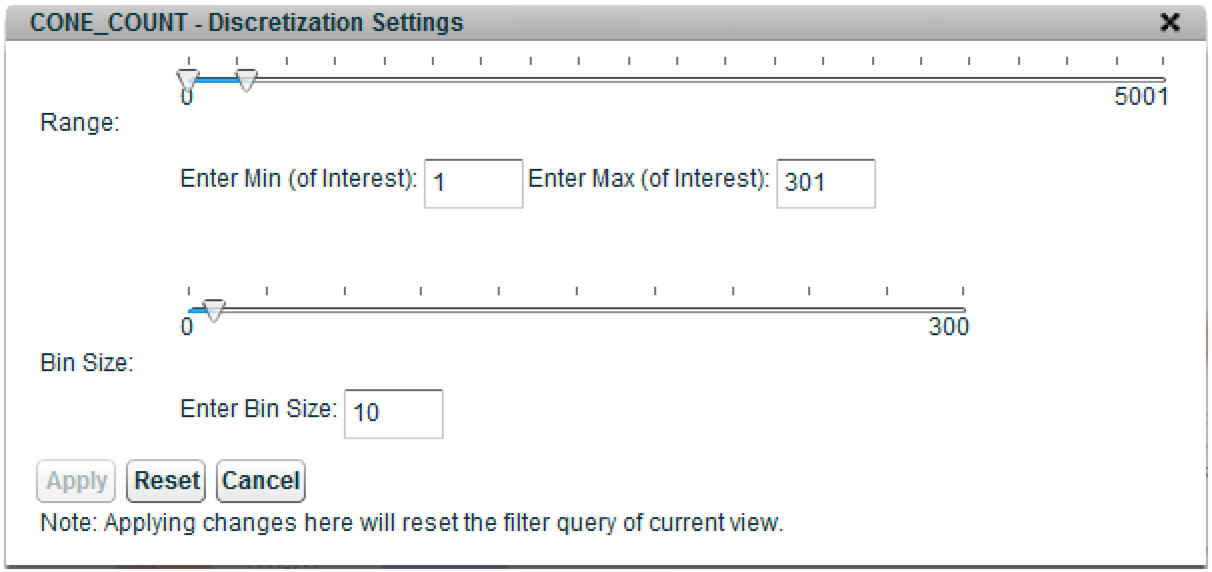


Line series of average cone production from 'Pinus' species (pine trees) from 1962-2011 showing a declining trend. Error bars are displayed. Users mouse over the line series for additional information.



The time series graph shows the very high degree of synchrony among cone crops of Abies amabilis (Pacific silver fir) in seven National Forests stretching from the Canadian border to the central western Cascades of Oregon, with an alternating 2- and 3-year cycle, and peak cone production in 1968. This graph suggests hypotheses about extrinsic factors, like climate, that may contribute to regional low cone production (e.g. in 1966-67, 1972-73, and 1983-84) and intrinsic factors, like energy expenditure in cone production, that may contribute to cycles.

Discretization setting



Discretization settings for attribute CONE_COUNT. Ecologists can narrow the range of interest for this attribute to [1, 301] and specify a bin size of 10. The result will automatically include two separate out-of-range buckets for [0, 1) and [301, 5001).

High-level Category	Task Type	DM's supported features	Specific analytical needs of target users
Data and View Specification	Visualize	Choose among small multiples views and time-series line charts	Inspect distributions of attributes with small multiples views and temporal trends with time-series line charts
	Filter	Filter data based on selection of buckets	Examine different data subsets or samples of observations
	Sort/ Reorder	Sort buckets within an attribute by names or by abundances	Organize the data according to a familiar unit of analysis (e.g., rank species from rare to common)
		Reorder attribute axes	Group axes by their common or user-defined characteristics (e.g., group of covariate/response variables)
	Derive	Discretize quantitative attributes	Experiment different features of the data with different discretization settings (e.g., isolate specific range of interest)
		Group/ungroup buckets within an attribute	Group outliers or similar attribute values to fit users' hypotheses (e.g., group species of the same genus or family)
		Scale (normalize) buckets' abundances	Accommodate data sets with different distributions
View Manipulation	Select/ Highlight	Select or highlight a view, axes, buckets, or line series	Select or highlight elements of interest for other operations, such as filter, sort, derive
	Navigate	Navigate and control views on canvas using the top menu bar and the bottom status bar	Know where and how to navigate views
	Coordinate	Duplicate small multiples views	Compare data subsets side-by-side
		Use a small multiples view as a visual query builder to construct series data for line charts	Construct multiple line series and compare them
	Organize	Open, close, resize, and layout views	Manage views for comparison or effective presentation to others
		Orient small multiple views vertical or horizontal	Accommodate users' orientation preference
		Show/hide labels in small multiple views Show/hide error bars in time-series line charts	Access additional information on demand
Process and Provenance	Record	Log user interactions	Undo/redo actions, reproduce states step-by-step. These features are reserved for future work.
	Annotate	Color axes and Label line series	Distinguish among axes or line series based on their common or user-defined characteristics
	Share	Bookmark states	Revisit/share states with others for collaborative and iterative exploration of data
		Export view data	Analyze data further with statistical tools.
	Guide	Display data tips for menu bars, axes, buckets, and line series	Guide users through menu items and provide additional information on highlighted items
		Enable rich data pop-ups for buckets or axes.	Display researcher-provided information potentially on any selected elements such as buckets or axes

Interaction features supported by the tool. Each of the features is designed to facilitate specific analytical needs of ecologists.

Vegetation Cover Data Set: 1979 - 2007

Vegetation observations:

- **159,658** observations
- **193** plots of **14** species
- **8** sampling years

Variable Name	Type	Description
SPECIES	nominal	Tree species code
PLOT	nominal	Plot number of 250 square meter plot
WATERSHED	nominal/ spatial	Watershed (WS01 or WS03)
UNIT	nominal	Watershed / Cutting unit number
PLANT_COMMUN	nominal	Plant community code
SOIL_TYPE	nominal	Soil type code
SOIL_DISTURB	nominal	Soil disturbance code
ELEV_LIDAR	quantitative	Elevation as determined from 1 meter LIDAR at GPS plot coordinates in NAD83
ASPECT	nominal	General aspect of plot by 8 cardinal directions
YY_MM	ordinal/ time-based	Year and month of measurement
BDIA	quantitative	Basal diameter (centimeters)
HDIA	quantitative	Diameter at breast height (centimeters)
VIGOR	nominal	Vigor of the tree
STATUS	nominal	The status of the tree


Firefox
Visual Analysis | Vegetation Data Set (tp0...
web.engr.oregonstate.edu/~pham/dm/veg/#
Data set: tp073
Settings
Selected Axes
Selected Bins
Current View: Main
Current State
Help
Main
PLOT
WATERSHED
UNIT
PLANT_COMMUN
SOIL_TYPE
SOIL_DISTURB
ELEV_LIDAR
ASPECT
YY_MM
SPECIES
BDIA
HDIA
VIGOR
STATUS
WS03
WS01
TSHE
THPL
TABR
RHPU
PSME
PREM
POTR2
LIDE2
CONU
CACH
ARME
ALRU
ACMA
ABAM
W
SW
SE
S
NW
NO DATE
NE
N
E
_NO_DATA
Line Chart: YY_MM - BDIA
BDIA_AVG
Line Chart: YY_MM - HDIA
HDIA_AVG
Line Chart: YY_MM - ELEV_LIDAR
ELEV_LIDAR_AVG
YY_MM
ACMA_ALL
CACH_ALL
PREM_ALL
PSME_ALL
THPL_ALL
TSHE_ALL
Filter Query: ALL
Inverse
Reset

Discussion: Data Analysis Approaches and Tools

What are **the approaches** you take to analyzing ecological long-term data sets?

What **tools** have you commonly used to analyze ecological long-term data sets?

Discussion: Explanatory Visualizations

What **types of charts** have you used for **presentation or explanation** of ecological long-term data?

What **types of knowledge or findings** have you presented or explained with these charts?

Discussion: Exploratory Visualizations

Have you used charts/graphs to **explore** ecological long-term data?

What **types of charts** have you used to explore ecological long-term data?

Are the charts you have used **interactive**?

What are the **interaction features** that you have commonly used?

What are the **interaction features** that you wish the charts support?

Sense-making approaches

Top down approach (Exploratory analysis)

Scientists who are interested in general ecological phenomena and may have little particular knowledge of the data.

They are in need of an exploratory tool that can help them formulate hypotheses and questions quickly following different exploration paths.

Bottom up approach (Confirmatory analysis)

Scientists who have collected the data and studied them intensively

Pre-establish specific hypotheses and questions on the data set

May skip the exploration step and use statistical tools to test these hypotheses directly.

What is Visualization?

Tamara Munzner 2011:

"Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively"